REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0397

ed
0188

Public reporting burden for this collection of information is estimated to average 1 hour per re ... sting data sources, gathering and maintaining the data needed, and completing and reviewing the collection of in ... other aspect of this collection of information, including suggestions for reducing this burden, to Washington Head ... rts, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management an ... DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED Final - 15 December 1994 - 14 June 1998

4. TITLE AND SUBTITLE The Pattern of Acoustic Cues Mediating Spatial Hearing Performance	5. FUNDING NUMBERS F49620-95-1-0106
6. AUTHOR(S) Rorbert H. Gilkey Department of Psychology	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Wright State University Dayton, OH 45435	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NL 801 North Randolph Street, Rm 732 Arlington, VA 22203-1977	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We have submitted, revised, and/or published 17 papers, chapters, and books. In addition we have made 17 presentations at various meetings. Some of the results reported in these papers ar based on the research efforts begun under AFOSR NL-91-0289 (including work on masked detection and sound localization in noise. We continue our theoretical work on spatial hearing. We have collected new data concerning the localization of speech stimuli, the effects of the listening envrionment on the perception of virtual audio and auditory-aided visual search. We have developed hardware and software to support planned experiments in a number of topic areas and we have published an edited book on binaural and spatial hearing in real and virtual

14. SUBJECT TERMS Acoustics, hearing	15. NUMBER OF PAGES 19
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclas	Unclas	Unclas	

Final Technical Report
December 15, 1994 to June 14, 1998

The Pattern of Acoustic Cues Mediating Spatial Hearing Performance
AFOSR F49620-95-1-0106

Robert H. Gilkey, PI
Department of Psychology
Wright State University
Dayton, OH 45435
(513) 873-3337
gilk@desire.wright.edu

I. OBJECTIVES

The long-term goals of the program of research, of which this project is a part, are to specify the mechanisms that underlie the spatial hearing abilities of humans and to apply this knowledge to applications such as auditory displays and virtual environment generation. The specific objectives were to answer a series of basic science questions concerning: the accuracy of sound localization judgments and the mechanisms that allow us to "hear out" and process one particular stimulus in the presence of other interfering stimuli. The results relate to a series of applied questions concerning the effectiveness of three-dimensional virtual auditory displays, when those displays are complex (e.g., containing many stimuli) or when they are used in a noisy environment (e.g., a cockpit). The results have also been used to develop and evaluate models of spatial hearing. In addition to these objectives, we have examined spatial hearing performance in rooms with reverberation, examined auditory determinants of the sense of "presence" in virtual environments, and prepared an edited book on binaural and spatial hearing in real and virtual environments.

II. STATUS OF EFFORT

We have submitted, revised, and/or published 17 papers, chapters, and books. In addition, we have made 17 presentations at various meetings. Some of the results reported in these papers are based on research efforts begun under AFOSR NL-91-0289 (including work on masked detection, Section III.A.1 and Section III.A.2; and sound localization in noise, Section III.B.1 and Section III.B.2); we continue our theoretical work on spatial hearing (Section III.E and Section III.F); we have collected new data concerning the localization of speech stimuli (Section III.B.3), the effects of the listening environment on the perception of virtual audio (Section III.C), and auditory-aided visual search (Section III.D); we have developed hardware and software to support planned experiments in a number of topic areas (Section III.C and Section III.H); and we have published an edited book on binaural and spatial hearing in real and virtual environments (Section III.G).

III. ACCOMPLISHMENTS

Much of the work described here was conducted in the Auditory Localization Facility of the Armstrong Laboratory at Wright-Patterson Air Force Base. This facility contains a 14-foot diameter geodesic sphere, with 277 speakers mounted on its surface. This is a unique facility that allows the experimenter considerable control over the spatial distribution of sound sources when conducting sound localization or free-field masking research. Additional studies are being performed in the Signal Detection Laboratory of the Department of Psychology at Wright State University. This is a more traditional psychoacoustic facility, where subjects listen to sounds presented over headphones in individual sound-attenuating booths. Many of the projects described here were begun with support from a previous AFOSR grant (NL-91-0289). Some of the work received additional support from Armstrong Laboratory, from a grant from the National Institutes of Health, from the Ohio Board of Regents, and through cost-sharing funds from Wright State University.

A. Masked Detection

1. Free-Field. Some of our work on free-field masking replicates previous work that has shown a substantial increase in detectability when the signal and masker are spatially separated [e.g., K. Saberi, L. Dostal, T. Sadralodabai, V. Bull, and D.R. Perrott, J. Acoust. Soc. Am. 90, 1355-1370 (1991)]. However, in our work the stimulus frequency was systematically manipulated. When the signal was separated from the masker in azimuth in the free field, the detectability of the signal could be increased by as much as 18 dB. Increases in detectability of as much as 8 dB were observed for separations in elevation. In all cases, the increases in detectability observed for our high-frequency (above 3.5 kHz) signal and masker were as great, or greater, than those observed for the low-frequency (below 1.4 kHz) signal and masker. Traditional models of binaural masking, based on interaural differences, did not predict the effects of stimulus frequency or the increases in detectability observed with vertical separations.

2. Virtual Sounds. The effects of spatial separations were compared for "real" and "virtual" sounds, in order to determine the relative importance of monaural and binaural cues for detection. Because the virtual stimuli were presented through headphones, monaural and binaural presentations could be compared by merely turning off one channel. Although there was some evidence suggesting a small role for interaural cues at low frequencies, in most cases the best monaural performance was as good as binaural performance, suggesting that the increases in detectability observed in the free field could have been mediated by monaural changes in the effective signal-to-noise ratio, rather than by changes in interaural information.

3. Publications. Three of our papers describe this work: Gilkey and Good (1995); Good, Gilkey, and Ball (1997); Gilkey, Good, and Ball (in revision).

B. Sound Localization in Anechoic Environments

1. Effects of Signal-to-Noise Ratio. In many situations, the observer must be able not only to detect a signal, but also to determine the direction of a sound source, once it has been detected. Because of the increased complexity of the localization task relative to the detection task, a more complete representation of the signal information is needed for accurate localization. In one experiment, the subject's task was to localize a click-train signal, which could originate from any of 239 directions surrounding the subject in azimuth and ranging from -45° to +90° in elevation. In some conditions a broadband Gaussian noise masker was presented from the speaker directly in front of the subject within the horizontal plane. Localization performance was measured in the quiet and at nine signal-to-noise ratios, ranging from -13 dB to +14 dB relative to the detection threshold for the signal when presented through the same speaker as the masker. The accuracy of the localization judgments decreased nearly monotonically as the signal-to-noise ratio was decreased. However, the accuracy of subjects' judgments relative to the frontal plane (the Front/Back dimension) was disrupted even at relatively high signal-to-noise ratios, but the accuracy of their judgments relative to the median plane (the Left/Right dimension) was not similarly disrupted, unless the signal-to-noise ratio is reduced considerably. There are important implications of these results for the design of auditory displays. Information about the laterality of the signal, whether it is to the left or to the right of the user, is likely to be faithfully represented even in adverse acoustic environments. However, we can anticipate that users will have difficulty determining whether the signal is in front of them or behind them when the signal-to-noise ratio is unfavorable. Elevation information, whether the signal is above or below the user, will not be transmitted as effectively as Left/Right information, but will in general be more reliable than Front/Back information.

2. Effects of Masker Location. In another experiment, the location of the masker was systematically varied. In different blocks of trials, the masker could be in front of the subject, behind the subject, directly to the subject's left, directly to the subject's right, or directly above the subject. At low signal-to-noise ratios, the subject's judgments of the direction of the signal were, in general, biased toward the direction of the masker. However, the location of the masker influenced this pattern of results in a complex manner. For some combinations of masker location, signal location, and signal-to-noise ratio, responses appeared to be biased away from the masker. Some masker locations appear to have a more general disruptive effect on localization performance (e.g., the masker location directly above the subject's head). Our examinations of the data from this experiment suggest that the pattern of results observed in the experiment described in Section III.B.1 was partially dependent on the location of the masker. Although performance in the Front/Back dimension is generally worse than in the other two dimensions, the decrease in performance as signal-to-noise ratio was lowered was most rapid when the masker was in front of the subject or behind the subject.

3. Localization of Complex Stimuli. Whereas most laboratory work on sound localization has used relatively simple stimuli with flat long-term power spectra (e.g., clicks or noise) most applications of spatial hearing technology are likely to use stimuli with more complex spectra. For example, one suggested application of spatial hearing technology is to add virtual spatial cues to a wingman's communication channel allowing the pilot to determine the location of the wingman's plane simply by monitoring the perceived location of his or her voice. A potential

problem arises because speech stimuli are likely to be more difficult to localize than stimuli with flat spectra; speech has comparatively little high-frequency energy and the shape of the speech spectrum varies from time to time, making it more difficult for the system to recover the spectral effects of the pinna. Previous studies with speech stimuli have considered the accuracy of subjects' azimuth judgments (which are likely to be based on low-frequency interaural difference cues), but have not systematically investigated the accuracy of elevation judgments (which are likely to be based on high-frequency spectral cues). In our study, subjects' accuracy with speech stimuli was comparable to that with click-train stimuli in the Left/Right dimension. However, judgments to click-train stimuli were consistently less accurate in the Front/Back dimension, and typically more accurate in the Up/Down dimension, than judgments to speech targets. These results indicate that localization performance in applied settings, using speech stimuli, may be less accurate than would be expected based on the bulk of the previous literature.

4. Publications and presentations. Several talks and papers describe this work, including: Gilkey (1995); Gilkey and Anderson (1995); Gilkey and Simpson (1996); Gilkey, Isabelle, Simpson, and Janko (1996); Gilkey, Simpson, Isabelle, and Anderson, and Good (1997); Good and Gilkey (1997); Good, Gilkey, and Ball (1997); Isabelle, Gilkey, Simpson, and Janko (1997); and Gilkey, Isabelle, and Simpson (1997a, 1997b).

C. Spatial Hearing in Reverberant Environments

The vast majority of data on spatial hearing has been acquired under anechoic conditions. However, the vast majority of acoustic environments are echoic. We have begun a series of experiments to evaluate localization performance in rooms.

Casual observations made while listening to binaural recordings suggest that virtual audio is "more compelling" when the listener hears the sounds in the same room where the original recordings were made. In a first attempt to quantify this effect, binaural recordings were made of "everyday" sounds such as keys jingling, a telephone ringing, speech, etc., in three different rooms, using the KEMAR manikin. The rooms ranged in volume from 16 m^3 to 194 m^3. All rooms were approximately square, and had hard walls and carpeted floors, such that the main difference between them was their size and reverberation time. During the experimental trials, a naive subject was seated in one of the three rooms with his/her head in the same position that KEMAR's head had been in when the recordings were made. The subject was given an opportunity to become familiar with the room, both auditorily and visually. During the experiment the subject listened through earphones to a recording made in a single one of the three rooms (not necessarily the same room as the room in which he/she was seated). On each trial, a single sound was presented and the subject's task was to indicate the perceived location of the sound, by making two marks (one indicating azimuth and distance, and one indicating elevation) on a response sheet that showed a graphical representation of the listening environment.

Localization errors were analyzed using the 3-pole coordinate system. Subjects were found to be most accurate in the left/right dimension, and less accurate in the front/back and up/down dimensions. This was consistent with results in the literature for localization of virtual stimuli. However, the overall magnitude of the errors in this experiment was larger. Contrary to our expectations, we found no significant differences in localization performance across conditions. That is, subjects' localization errors did not appear to be systematically affected by the listening room or by the recorded room.

We also developed a questionnaire that was completed by each subject at the end of an experimental session. The questionnaire was designed to measure the degree to which each subject experienced a sense of presence in the auditory virtual environment. Results from the questionnaire show that subjects experienced a greater sense of presence when the listening room and recorded room were the same, suggesting that the sense of presence is indeed affected by the listening environment.

Although these indications of greater presence were statistically significant, the effects were small. Therefore, we attempted to develop a more sensitive measure of presence. Subjects again sat in a single one of the three rooms during the experiment. On each trial, two virtual stimuli were

presented in sequence, each recorded in a different room, and subjects were asked to indicate which stimulus was more realistic. Figure 1 shows the percent preference, averaged across all subjects, for a particular sound as a function of listening room and recorded room reverberation time. The columns with the cross-hatched tops indicate cases in which the listening room and recorded room were matched. We found that a subject was more likely to choose a sound recorded in particular room as the most realistic when the subject was listening in that room.

We have planned a series of experiments to study further the relative contribution of non-auditory factors in achieving a sense of presence in auditory virtual environments. For example, we will use the VERITAS facility (see section III.H.1) to present virtual audio while the subject views the visual objects that are potential sources of the sounds. We are constructing high-resolution near-photorealistic virtual models of the visual aspects of the three rooms used by Simpson et al. We plan to have subjects listen to the same virtual sounds used in the Simpson et al. study while they are seated in the CAVE™ and viewing a visual representation of one of the three rooms. If the visual representation of the virtual room has a similar influence on the subjects as the real room and if the effects observed by Simpson et al. were mediate by the subjects' visual experience with the "real" room, then we should observe similar effects in both experiments. If so, this will provide evidence that the positive effect of auditory stimulation on the sense of presence is dependent on the "match" between auditory and visual stimulation (a situation that is at best partially realized in typical virtual environments).

This work is described in Gilkey, Isabelle, and Simpson (1997a, 1997b); Simpson (1997); and Simpson, Hale, Isabelle, and Gilkey (1996).

D. Auditory-aided visual search

One of the promising applications of 3-D auditory displays is to direct the attention of the user toward relevant information. For example, spatialized sound could be used to direct a pilot's attention to an important visually displayed instrument or to a potential threat outside the cockpit. Previous research in the Armstrong Laboratory by Perrott et al. [Human Factors Proceedings, 104-108 (1995)] shows that search times for an isolated light against a dark background could be reduced by 10-50% when an auditory cue was present. Results from some experiments suggest that the greatest benefits of spatialized auditory cues are seen when the visual search task is perceived to be the most complex [Nelson et al., in press]. In a recent experiment conducted in our laboratory, using a more difficult visual search task, we found a much larger effect of the auditory cue. In our experiment, the subject wore a head-mounted display (HMD) with a limited field-of-view (40° horizontal by 20° vertical), and looked at a virtual array of letters that surrounded him/her in azimuth, and ranged from -30° elevation to +30° elevation. All of the letters, except the target, were either "capital Ps" or "capital Qs." The subject's task was to find the single "capital R" (i.e., the target). Characters were positioned in 5° by 4° grid cells, such that approximately 40 characters were visible in the HMD at any time. The subject searched the entire field of letters until the "R" was found. In the auditory-aided condition, a virtual auditory cue (filtered with head-related transfer functions, presented through headphones, and fixed in virtual space using a head tracker) was presented near the virtual spatial location of the target. Figure 2 shows the average target acquisition time across 5 subjects for the visual only and auditory-aided visual search conditions. Acquisition times decreased by more than a factor of 8 when the auditory cue was added. Note also that this increase in speed was realized with a relatively poor auditory display (i.e., non-individualized head-related transfer functions, no reverberation model, and no interpolation between recorded spatial locations such that the auditory signal could be as much as 9° away from the center of the visual target).

In future research, we plan to map the relations between auditory and visual mechanisms for search, with particular emphasis on how the relative quality of auditory and visual displays trade off to determine the utility of spatialized auditory cues. For example, a chromatic visual target will be presented in a relatively high-density field of white distracters. Visual search times will be longer when the target chromaticity is low (i.e., the target appears more similar to the distracters), than when the target chromaticity is high. We hypothesize that the auditory cue will lead to the greatest

reduction in search times for such conditions. Some of these experiments will utilize the VERITAS facility (see section III.H.1).

These results have been described in Gilkey (1996); Gilkey, Isabelle, Simpson, and Janko (1996), Isabelle, Gilkey, Janko, and Simpson (1997); Gilkey, Isabelle, and Simpson (1997a, 1997b); Gilkey and Simpson (1996); and Gilkey, Simpson, Isabelle, Anderson, and Good (1997).

E. Neural Network Models of Sound Localization

1. Localization in the quiet. At least three types of acoustic cues are generally recognized as providing the foundation for sound localization: interaural time differences, interaural level differences, and direction-specific spectral modulations introduced by the acoustics of the torso, head, and pinnae. No model has been developed to describe how these disparate sources of information are combined into a single unified perception of the source location. Because sound localization can be seen as requiring the listener to associate the pattern of acoustic cues received on a given trial with a particular source location and because neural networks have had great success in solving other pattern recognition problems, we have been using them to model sound localization.

Our initial models were composed of a preprocessing stage and a neural network stage. In the preprocessing stage, the click signals were convolved with head-related transfer functions (filters that simulate the acoustic effect of the torso, head, and pinnae) and corrupted by internal noise. The interaural delay corresponding to the maximum in the cross-correlation function between the noisy waveforms in the left and right channels was used as one possible input to the neural-network section of the model. In addition, the energy in each of 22 rectangular quarter-octave bands was determined for both the left and right channels. Logarithms of these quarter-octave spectra, or the difference between the log spectra in the left and right ears, were also possible inputs to the neural-network stage.

Although several configurations of the neural-network stage have been considered, two are of particular interest. Because there has been some controversy in the literature as to whether spectral information used for sound localization is represented in the system via monaural processing or via interaural processing, we configured one network to utilize monaural spectral information, and another to utilize interaural spectral information. Performance for the interaural network (which received the interaural difference spectrum and the interaural time difference) was slightly better than that of the human subject, whose head-related transfer functions were used in the pre-processing stage (this subject is generally recognized as a good localizer) in the Left/Right, Front/Back, and Up/Down dimensions. Thus, these results indicated that there is sufficient information in the binaural representation of the stimulus to mediate human-like sound localization performance.

A monaural model can also achieve performance similar to human performance. We first trained separate networks to localize based on the spectrum in the left ear and based on the spectrum in the right ear. Performance for either of these monaural networks was, in at least some situations, notably worse than human performance, and showed an asymmetry, with better performance seen on the ipsilateral side of the head (a pattern not evident in the human data). We, therefore, used the outputs of both of these networks as inputs to a third, arbitrator, network. This hierarchical network did not show an asymmetry and performed nearly as well as the human subject in the Left/Right dimension, but somewhat worse than the human subject in the Front/Back and Up/Down dimension. In this case, binaural interaction, in the traditional sense, was not possible because the arbitrator network combined the "decisions" from the left and right channel rather than the stimuli. It could be argued that such a "pure" monaural model is a bit of a straw man. No one would argue that normal human sound localization occurs without the use of interaural time differences. Thus, by providing the interaural time delay as an additional input to the arbitrator network, we create a model that has "normal" interaural timing information to determine the left/right dimension, but does not have normal interaural level information for determining front/back and up/down dimensions. Despite this, performance comparable to the human was observed in all three dimensions, indicating that up/down and front/back performance could be based on monaural processing alone.

Overall, these analyses indicate that either monaural or interaural spectral information, in combination with the interaural time delay, is sufficient to mediate human-like sound localization performance.

2. Localization in noise. Because we were able to predict sound localization in quiet using two quite different models (i.e., one based on a binaural representation of the "pinna effects" and one based on a monaural representation), we are using our data on localization in noise to further constrain the form of the models. However, in order to represent the effects of external noise in a nontrivial manner, a more detailed model of the auditory periphery and binaural interaction is needed. Specifically, we are investigating a more-realistic model for the human auditory periphery consisting of: a gammatone filter bank (to model cochlear frequency selectivity), followed by ν-law rectification (to model haircell transduction), and lowpass filtering (to model frequency-dependent phase-locking in auditory nerve fibers). This model captures the first-order spectral and temporal aspects of the auditory periphery; however, some phenomena, such as amplitude-dependent bandwidth changes, are neglected in the interest of computational efficiency. We have investigated two representations of binaural interaction: a traditional crosscorrelator model and a crosscorrelator with inhibition. In the latter, the binaural interaction has the form of running interaural cross-correlation with inhibition for each frequency channel, similar to that of Lindemann [J. Acoust. Soc. Am. 80, 1608-1622 (1986)] such that earlier-arriving signals from one ear attenuate later-arriving signals from the opposite side. One major effect of the inhibition is to enhance contrast in the "cross-correlation" pattern ("sharpening of the peaks").

Binaural models often display information in two dimensions; as a function of both the correlation lag (t) and the center frequency (f) of the peripheral bandpass filter. It has been argued that consistency in ITD across frequency (i.e., straightness in the τ/f representation) is an important aspect for localization.

We have trained neural networks using the τ/f representation, averaged across running time, and have found that while that model performance on the dimensions of the three-pole coordinate system, left/right (L/R) , up/down (U/D), and front/back (F/B), is ordered similarly to human performance (e.g., L/R is best, followed by U/D, with worst performance in the F/B dimension), for the case when we set the internal noise level to best predict human model performance in the quiet. However, the model does not do a good job predicting the change in human performance as a function of external noise level. Specifically, when a net is trained at a given external noise level and then tested at that same level, it tends to show better performance than humans in all three dimensions (L/R, U/D, F/B). However, a net that is trained at a given external noise level, but then tested at a different level of noise, will show worse than human performance, even if the noise level used during testing is less than the noise level used during training (i.e., tested with a more-favorable signal-to-noise ratio than trained on).

We have concluded that the set of features encoded by the neural nets when trained in this manner is dependent on the signal-to-noise ratio in a way that is incompatible with human performance. Specifically, we have seen that the peaks in the inhibited crosscorrelation pattern do not simply become less well-defined with increasing external noise, rather they shift in location as well. In Figure 2, Panel A shows the output from a single low-frequency channel from the inhibited crosscorrelation mechanism to a click-train target in the quiet, plotted for locations varying in azimuth from -180° to +180°. The location of the peak in the pattern varies systematically with the azimuth of the sound source. Panel B shows the same pattern for the case of a speech target in quiet (the word "pass" spoken by a male talker). In contrast, Panel C shows the pattern for the case of the click train target with higher levels of external noise (-10 dB SNR). Note that the peaks are still somewhat defined, but do not appear to vary systematically with the location of the sound source. Note that in this case the masker location is always at 0° azimuth corresponding to τ=0, and while the pattern is more clustered around τ=0, there are peaks at other values also. In contrast, we have observed that human performance, particularly in the L/R dimension, is still reasonably good at this level of external noise (Good and Gilkey, 1996b).

To explore this issue further, we have trained neural networks on a range of external noise levels. The model of binaural interaction was based on the inhibited cross-correlation model of Lindemann with no dynamic inhibition, C_d=0, and increased static inhibition, C_s=0.8. Within each frequency channel, the peripherally processed signals were used to compute the running-time inhibited interaural cross-correlation function, with correlation lags between ±1 ms. The resulting pattern was averaged across running time. The set of averaged cross-correlation patterns (one pattern for each frequency channel) was sampled at a 12.5-kHz rate and corrupted by uniformly distributed internal noise to provide the input to the neural network (for computational efficiency, no peripheral internal noise was added). The level of internal noise was adjusted so that the localization performance of the model, when trained and tested in the quiet, was comparable to that of a human observer in the quiet. However, the network was trained across multiple signal-to-noise (i.e., within the same training regiment, SNRs of 1 dB, 11 dB, 21 dB, and Quiet were presented; the internal noise level was held constant). Figure 3 plots rms localization error as a function of SNR. Panel A shows the average results for the three subjects of Good and Gilkey (1996a), and Panel B shows the results for the model, with separate functions for each of the three dimensions (L/R, F/B, and U/D). As can be seen in the figure, the functions for human and model are similar in terms of the ordering and slope, but are different in detail. In general, the model shows worse performance than humans, and in particular, the F/B error of the model is much larger at high SNRs. In addition, the error in quiet is larger for the model, which may indicate an excessive level of internal noise. (Recall, the level of internal noise was chosen for a network that was trained and tested in the quiet only. We expect that a lower level of internal noise would translate all points vertically, to lower levels of rms error.) A more detailed examination of the trial-by-trial responses of the model shows biases toward the masker location in the L/R and F/B dimensions like those observed for humans, but with a bias toward lower elevations than that of the masker in the U/D dimension. Although the pattern of front-back and back-front reversals changes with SNR in a manner similar to humans, the model also exhibits left-right reversals that are typically not observed for humans.

We consider these results (as reported in Isabelle, Janko, and Gilkey, 1998a and1998b) to be preliminary, but they indicate that this type of model may be able to predict the localization in noise data.

3. Publications and presentations. This work is described in Janko, Anderson, and Gilkey (1996); and Gilkey, Isabelle, Janko, and Simpson (1996, 1997), and Isabelle, Janko, and Gilkey (1998a,1998b).

F. The Role of Auditory Stimulation in Achieving a Sense of Presence in Virtual Environments

Ramsdell ["The psychology of the hard-of-hearing and the deafened adult," in Hearing and Deafness, edited by S.R. Silverman and H. Davis (Holt, Rinehart, and Winston, New York), 499-510 (1978)] reports that adventitiously-deafened individuals feel a sense of unconnectedness with their surroundings, a sense that the world seems "dead." Such reports offer a compelling rationale for the argument that auditory cues are a crucial determinant of the sense of presence. Moreover, the crucial element of auditory stimulation for creating a sense of "presence" may be the auditory background, comprising the incidental sounds made by objects in the environment, rather than the communication and warning signals that typically capture our attention. Although designers of virtual environments have most often tried to maximize the sense of presence in the user by attempting to improve the fidelity of visual displays, we argue that background auditory stimulation may be useful or even critical for achieving a full sense of presence.

A paper presenting this argument has been published: Gilkey and Weisenberger (1995).

G. Book on Binaural and Spatial Hearing in Real and Virtual Environments

The Conference on Binaural and Spatial Hearing was held at the Hope Hotel and Conference Center at Wright-Patterson Air Force Base, Ohio, on September 9-12, 1993 with AFOSR and Armstrong Laboratory as sponsors. We have compiled and edited a book entitled "Binaural and Spatial Hearing in Real and Virtual Environments," loosely based on the

conference. The book is intended to be more than a simple proceedings; the 34 chapters provide broad coverage of binaural and spatial hearing including: Chapter 1. Factors affecting the relative salience of sound localization cues (Wightman and Kistler); Chapter 2. Acoustical features of the human external ear (Shaw); Chapter 3. Elevation dependence of the interaural transfer function (Duda); Chapter 4. Spectral shape cues for sound localization (Middlebrooks); Chapter 5. Spatial referents of stimulus frequencies: Their role in sound localization (Butler); Chapter 6. Detection and discrimination of interaural disparities: Modern earphone-based studies (Bernstein); Chapter 7. Recent experiments concerning the relative potency and interaction of interaural disparities (Buell and Trahiotis); Chapter 8. The relative contributions of targets and distracters in judgments of laterality based on interaural differences of level (Dye); Chapter 9. Binaural masking level differences in nonsimultaneous masking (Kohlrausch and Fassel); Chapter 10. Listening in a room and the precedence effect (Hartmann); Chapter 11. Binaural adaptation and the effectiveness of a stimulus beyond its onset (Hafter); Chapter 12. The precedence effect: Beyond echo suppression (Clifton and Freyman); Chapter 13. Phenomenal geometry and the measurement of perceived auditory distance (Mershon); Chapter 14. Some observations regarding motion-without-direction (Perrott and Strybel); Chapter 15. Auditory motion perception: Snapshots re-visited (Grantham); Chapter 16. Experiments on auditory motion discrimination (Saberi and Hafter); Chapter 17. The cocktail party problem: Forty years later (Yost); Chapter 18. The relation between detection in noise and localization in noise in the free field (Good et al.); Chapter 19. Directional cueing effects in auditory recognition (Doll and Hanna); Chapter 20. Neural processing of binaural temporal cues (Kuwada et al.); Chapter 21. Neuronal processing for coding interaural time disparities (Yin et al.); Chapter 22. Auditory cortex and spatial hearing (Brugge et al.); Chapter 23. Head-related transfer functions in cat: Neural representation and the effects of pinna movement (Young et al.); Chapter 24. Models of binaural perception (Stern and Trahiotis); Chapter 25. Modeling binaural detection performance for individual masker waveforms (Colburn et al.); Chapter 26. Using neural networks to evaluate the viability of monaural and interaural cues for sound localization (Janko et al.); Chapter 27. Development of binaural and spatial hearing in infants and children (Litovsky and Ashmead); Chapter 28. An introduction to binaural technology (Blauert); Chapter 29. Auditory displays (Shinn-Cunningham et al.); Chapter 30. Binaural measurements and applications (Burkhard); Chapter 31. Flight demonstration of a 3-D auditory display (McKinley and Ericson); Chapter 32. The intelligibility of multiple talkers separated spatially in noise (Ericson and McKinley); Chapter 33. Binaural performance in listeners with impaired hearing: Aided and unaided results (Koehnke and Besing); Chapter 34. Signal processing for hearing aids employing binaural cues (Kollmeier). Several of these chapters provide extensive bibliographies. We anticipate that the book will be an important and widely used reference, both for hearing researchers and for scientists and engineers interested in the auditory component of virtual environment generation. The book was published in January of 1997 (Gilkey and Anderson, 1997).

H. Laboratory Development

1. <u>Virtual Environment Research, Interactive Technology, And Simulation (VERITAS) facility.</u> Our current research focus is shifting from solely auditory processing and auditory displays to multisensory displays and virtual environments. As indicated in sections III.D. and III.F., we are particularly interested in auditory-aided visual search and auditory-visual interactions in determining the sense of presence in virtual environments. To support this research, we have worked to establish a facility for virtual environment research. We received initial capital funding from the Ohio Board of Regents to establish the Virtual Environment Research, Interactive Technology, And Simulation (VERITAS) facility, which is owned and operated by Wright State University but housed in AL/CFBA at Wright-Patterson AFB.

VERITAS currently comprises a highly immersive visual display subsystem, and an integrated spatialized auditory display subsystem. The visual display subsystem consists of a CAVE™ (CAVE Automatic Virtual Environment), essentially a set of four rear-projection screens forming a cubical room, about 3.3 m on each side. High-resolution stereoscopic images are

displayed on the four walls and top-projected onto the floor by five CRT projectors (Marquee 8500, Electrohome). The user is nearly completely immersed, surrounded on all sides and from below with interactive stereoscopic images. The stereo field-sequential technique is used in which the user wears LCD shutter glasses (CrystalEyes, Stereographics), which synchronously block image transmission to one eye while the image for the unblocked eye is drawn on the screen, with right/left eye fields alternating at a rate of 120 Hz. The users field of view is limited only by the frames of the shutter glasses, which (similar to conventional eyeglasses) provide about 105° horizontal field of view. Because of the highly immersive surrounding display, 3-D virtual objects appear to fill the room. An artist's rendering of the CAVE is shown in Figure 4.

Imagery on the CAVE walls is generated by a Silicon Graphics (SGI) Onyx. The initial hardware configuration of the SGI Onyx includes: four R4400 CPUs; 256 Mbytes memory; 10 GB disk; one Infinite Reality graphics subsystem (including four Raster Managers and eight channels of video output); and three RS-232 serial ports. A 6DOF magnetic tracker (Flock of Birds, Ascension) is used to monitor the users head position and orientation in order to properly compute viewing perspective as the user moves about the entire interior of the CAVE. The users hand position and orientation is also magnetically tracked to provide a means for gestural control and interaction with virtual objects.

The spatialized auditory display subsystem (PowerSDAC, Tucker-Davis Technologies) provides 3D sounds over headphones. The users head position and orientation obtained from the magnetic tracker is also used to compute the appropriate acoustic perspective to the simulated sound sources. A network connection is used to transmit data on the users head position and orientation, and on the position and movement of virtual objects associated with virtual sounds, from the SGI Onyx to the host computer of the audio digital signal processing (DSP) engine, thereby maintaining synchronization between the visual and auditory attributes of virtual objects.

Our software orientation has been to purchase off-the-shelf software that will allow relatively inexperienced programmers to manipulate virtual environment generation. The virtual environment generation software used in VERITAS, Vega (Paradigm Simulation, Inc.), is built on the SGI Performer real-time 3D rendering library for optimal performance. It simultaneously provides both high-level programming constructs and a graphical user interface to reduce development time for sophisticated visual simulations. Vega provides us with wide-ranging choices from a number of third-party vendors that provide Vega-compatible solutions for other simulation needs (e.g., flight dynamics).

We are currently using additional funds from DURIP (#F49620-97-1-0118) and AFOSR (#F49620-97-1-0231) to enhance the capabilities of the VERITAS and to support our work on interface designs for Uninhabited Aerial Vehicles.

The VERITAS facility is described in: Isabelle, Gilkey, Kenyon, Valentino, Flach, Spenny, & Anderson, (1997a, 1997b); and Gilkey, Isabelle, & Simpson (1997a, 1997b).

2. Laboratory Move. During the Spring of 1996 the Signal Detection Laboratory was moved from Oelman Hall to Fawcett Hall.

3. Speaker Equalization. In order to provide the necessary spectral control of stimuli required in free-field localization and masked detection experiments, an equalization filter is designed for each loudspeaker, so that the effective stimulus at the source is the same independent of the speaker of origin. Our previous method of characterizing the loudspeakers in the Auditory Localization Facility employed repeated presentations of relatively long-duration wideband noise. We have developed a much faster method using pseudo-random pulse trains (e.g., Golay sequences), which makes it feasible (in terms of time and labor) to equalize the loudspeakers immediately prior to an experimental session, thereby taking into account the current effects of temperature and humidity. We can now implement these loudspeaker equalization filters using the Tucker-Davis Technologies PowerSDAC (a specialized high-speed digital signal processing system).

4. Reducing Incidental Echoes. The spherical array of loudspeakers in the Auditory Localization Facility contains many surfaces that introduce acoustic reflections (echoes) in the interior listening area. We have been investigating the nature of these echoes, with particular

concern regarding our planned experiments on localization in synthesized reverberant environments (i.e., we want only the echoes we intend to generate to be present in the environment). Simple models based on geometrical acoustics predict that only the loudspeaker directly opposite the activated loudspeaker contributes to a reflection. Our measurements have shown that reflections come from many loudspeakers on the hemisphere opposite an activated loudspeaker. We have been investigating both physical and signal processing approaches to reducing the level of the echoes.

5. Head-Related Transfer Function Measurement. We have adapted our loudspeaker equalization techniques to permit the measurement of head-related transfer functions, which capture the directionally dependent acoustic filtering of the torso, head, and pinnae. Individualized head-related transfer function recordings are often thought to be essential for perceptually adequate synthesized auditory displays. Our method uses time-domain techniques, which are required in order to compensate properly for the echoes in the sphere. Previous methods used frequency-domain techniques that result in head-related transfer function measurements contaminated by echoes. Further, our method results in a reduction of measurement time (to 6 minutes) over previous methods used in the Auditory Localization Facility (requiring 3 to 24 hours), making it feasible to acquire head-related transfer functions from live human subjects.

6. Binaural Room Impulse Response Measurement. We have extended our head-related transfer function measurement methods to record binaural impulse responses in environments with acoustic reflections and reverberation. The time-domain methods we use permit us to efficiently capture the temporal structure of the room impulse response, using binary sequence signals and signal processing equivalent to averaging responses to a large number of clicks but in a much shorter time. We have used these measurement techniques to analyze rooms we employ in the experiments described in section III.C.

7. Response Technique. In support of our localization research, a new pointing response technique was developed with the support of AFOSR NL-91-0289. A paper describing this technique was published during the period covered by this progress report: Gilkey, Good, Ericson, Brinkman, and Stewart (1995).

IV. KEY PERSONNEL

Receiving salary support:

Robert H. Gilkey, Ph.D., Associate Professor, Department of Psychology, Wright State University and Adjunct Research Scientist, AL/CFBA, Wright-Patterson AFB.

Scott K. Isabelle, Ph.D., Research Assistant Professor, Department of Psychology, Wright State University

Jennifer M. Ball, Graduate Research Assistant, Department of Psychology, Wright State University.

Brian D. Simpson, Graduate Research Assistant, Department of Psychology, Wright State University.

James Kondash, Graduate Research Assistant, Department of Psychology, Wright State University.

Jeffrey Shapiro, Graduate Research Assistant, Department of Psychology, Wright State University.

<u>James A. Janko</u>, Research Programmer, Department of Psychology, Wright State University.

<u>John M. Stewart</u>, Research Programmer, Department of Psychology, Wright State University.

<u>Affiliated</u>:

<u>Timothy R. Anderson</u>, Ph.D., Electronics Engineer, AL/CFBA, Wright-Patterson AFB.

V. PUBLICATIONS

<u>Papers Published</u>

Isabelle, S.K., Janko, J.A., and Gilkey, R.H. (1998a). Model of auditory localization in noise using neural networks. Proceedings of the 16th International Congress on Acoustics and 135th meeting of the Acoustical Society of America, Vol. 2, pp. 855-856.

Gilkey, R. H., Isabelle, S. K., & Simpson, B. B. (1997a). Virtual displays and virtual environments. <u>Journal or Ergonomics Society of Korea, 16,</u> 101-122.

Flach, J.M. & Gilkey, R.H. (1997a). Synthetic task environments and situation awareness. <u>Proceedings of the IEEE Aerospace and Electronic Systems Society 14th Annual Meeting: Synthetic Visualization: Systems and Applications</u>, 11-17.

Gilkey, R.H., & Anderson, T.R. (Eds.). (1997). <u>Binaural and Spatial Hearing</u>. Hillsdale, NJ: Erlbaum.

Gilkey, R.H., Isabelle, S.K., Janko, J.A., & Simpson, B.D (1997). Effects of uncertainty and masking on sound localization. In A. Schick (Ed.), <u>Proceedings of the 7th Oldenburg Symposium on Psychological Acoustics</u>.

Gilkey, R.H., Simpson, B.D., Isabelle, S.K., Anderson, T.A., & Good, M.D. (1997). Design considerations for 3-D auditory displays in cockpits. <u>Proceedings of the AGARD Conference on the Effectiveness of Auditory Displays in Aviation</u>, Copenhagen, Denmark.

Good, M.D., Gilkey, R.H., & Ball, J.M. (1997). The relation between detection in noise and localization in noise in the free field. In R.H. Gilkey & T.R. Anderson (Eds.), <u>Binaural and Spatial Hearing</u>. Hillsdale, NJ: Erlbaum.

Isabelle, S.K., Gilkey, R.H., Kenyon, R.V., Valentino, G., Flach, J.M., Spenny, C.H., & Anderson, T.R. (1997). Defense applications of the CAVE (CAVE Automatic Virtual Environment). <u>Proceedings of SPIE, 3057</u>, 118-125.

Janko, J.A., Anderson, T.R., & Gilkey, R.H. (1997). Neural network models of monaural and binaural sound localization. In R.H. Gilkey & T.R. Anderson (Eds.), <u>Binaural and Spatial Hearing</u>. Hillsdale, NJ: Erlbaum.

Good, M.D., & Gilkey, R.H. (1996a). Sound localization in noise: I. Effects of signal-to-noise ratio. <u>Journal of the Acoustical Society of America, 99,</u> 1108-1117.

Gilkey, R.H. (1995). Some considerations for the design of auditory displays. Proceedings of the IEEE Workshop on Applications of Signal Processing in Audio and Acoustics, (IEEE, New York), Paper 4.2.

Gilkey, R.H., & Anderson, T.R. (1995). The accuracy of absolute localization judgments for speech stimuli. Journal of Vestibular Research, 5, 487-497.

Gilkey, R.H., & Good, M.D. (1995). Effects of frequency on free-field masking. Human Factors, 37, 835-843.

Gilkey, R.H., Good, M.D., Ericson, M.A., Brinkman, J., & Stewart, J.M. (1995). A pointing technique for rapidly collecting localization responses in auditory research. Behavior Research Methods, Instrumentation, and Computers, 27, 1-11.

Gilkey, R.H., & Weisenberger, J.M. (1995). The sense of presence for the suddenly-deafened adult: Implications for virtual environments. Presence: Teleoperators and Virtual Environments, 4, 357-363.

Papers submitted:

Gilkey, R.H., Good, M.D., & Ball, J.M.. A comparison of "free-field" masking for real and for virtual sounds. Journal of the Acoustical Society of America, in revision.

Isabelle, S.K., and Colburn, H.S. Models of binaural detection: Comparisons to new experiments using reproducible noise. *J. Acous. Soc. Am.* (in revision).

Papers in preparation:

Meyer, T.A., & Gilkey, R.H. (1997). Models of subject response in a reproducible-noise masking task: Across-channel listening. Journal of the Acoustical Society of America, in preparation.

VI. INTERACTIONS

Presentations at meetings

Isabelle, S. K., Janko, J.A. & Gilkey, R. H. (1998b). A model of auditory localization in noise using neural networks. Journal of the Acoustical Society of America, 103, 2845

Simpson, B.D., Isabelle, S.K., and Gilkey, R.H. (1998). "The Sense of Presence in Auditory Virtual Environments," poster presented at the Conference on Human Interaction in Complex Systems. Dayton, OH.

Gilkey, R.H., Isabelle, S.K., & Flach, J.M. (1997). Cognitive workload in a complex synthetic task environment: The role of functional display and control representations. New World Vistas "People" Program Kickoff, (San Antonio, TX, May).

Isabelle, S.K., Gilkey, R.H., Kenyon, R.V., Valentino, G., Flach, J.M., Spenny, C.H., & Anderson, T.R. (1996). Defense applications of the CAVE (CAVE Automatic Virtual Environment). SPIE AeroSense97: Flat panel displays in cockpits, (Orlando, FL, April).

Gilkey, R.H., Isabelle, S.K., & Simpson, B.D. (1997b). Spatial hearing and virtual environments. Keynote address at the Meeting of the Korean Ergonomics Society, (Seoul, Korea, April).

Flach, J.M. & Gilkey, R.H. (1997b). Synthetic task environments and situation awareness, IEEE:AESS 14th Annual Meeting: Synthetic Visualization: Systems and Applications, (Dayton, OH, April).

Simpson, B.D., Hale, D.W., Isabelle, S.K., & Gilkey, R.H. (1996). The experience of naive subjects listening to virtual sounds, J. Acoust. Soc. Am. 100, 2633 (A).

Gilkey, R. H. & Simpson, B.D. (1996). Design considerations for 3-D auditory displays in cockpits. AMP Symposium on Audio Effectiveness in Aviation, (Copenhagen, Denmark, October).

Gilkey, R.H., Isabelle, S.K., Simpson, B.D., & Janko, J.A. (1996). Effects of uncertainty and masking on sound localization. 7th Oldenburg Symposium on Psychological Acoustics and Summer Course of the Graduate College "Psychoacoustics", (Oldenburg, Germany, August), Paper 6.1.

Gilkey, R.H. (1996). Issues in the implementation of 3-dimensional audio displays. AFOSR Review of Basic Research in Human Vision, (WPAFB, Ohio, June).

Gilkey, R.H. (1996). Some considerations for the design of auditory displays. 2nd International Workshop on 3-D Imaging Media Technology, (Seoul, Korea, March).

Gilkey, R.H. (1995). Some considerations for the design of auditory displays. IEEE Workshop on Applications of Signal Processing in Audio and Acoustics, (New Paltz, New York, October).

Other talks

Simpson, B.D. (1997). A sense of presence in auditory virtual environments. Brown Bag Lecture Series, Department of Psychology, Wright State University, (Dayton, OH, January).

Isabelle, S.K., Janko, J.A., & Gilkey, R.H. (1996). On the modeling of auditory localization of multiple sources. Hearing Research Seminar, Department of Biomedical Engineering, Boston University, (Boston, MA, November).

Gilkey, R.H. (1996). Some considerations for the design of auditory displays. Presented at the Korea Research Institute of Standards and Sciences (Taejon, Korea, March).

VII. PATENTS
None.

VIII. HONORS/AWARDS
None.

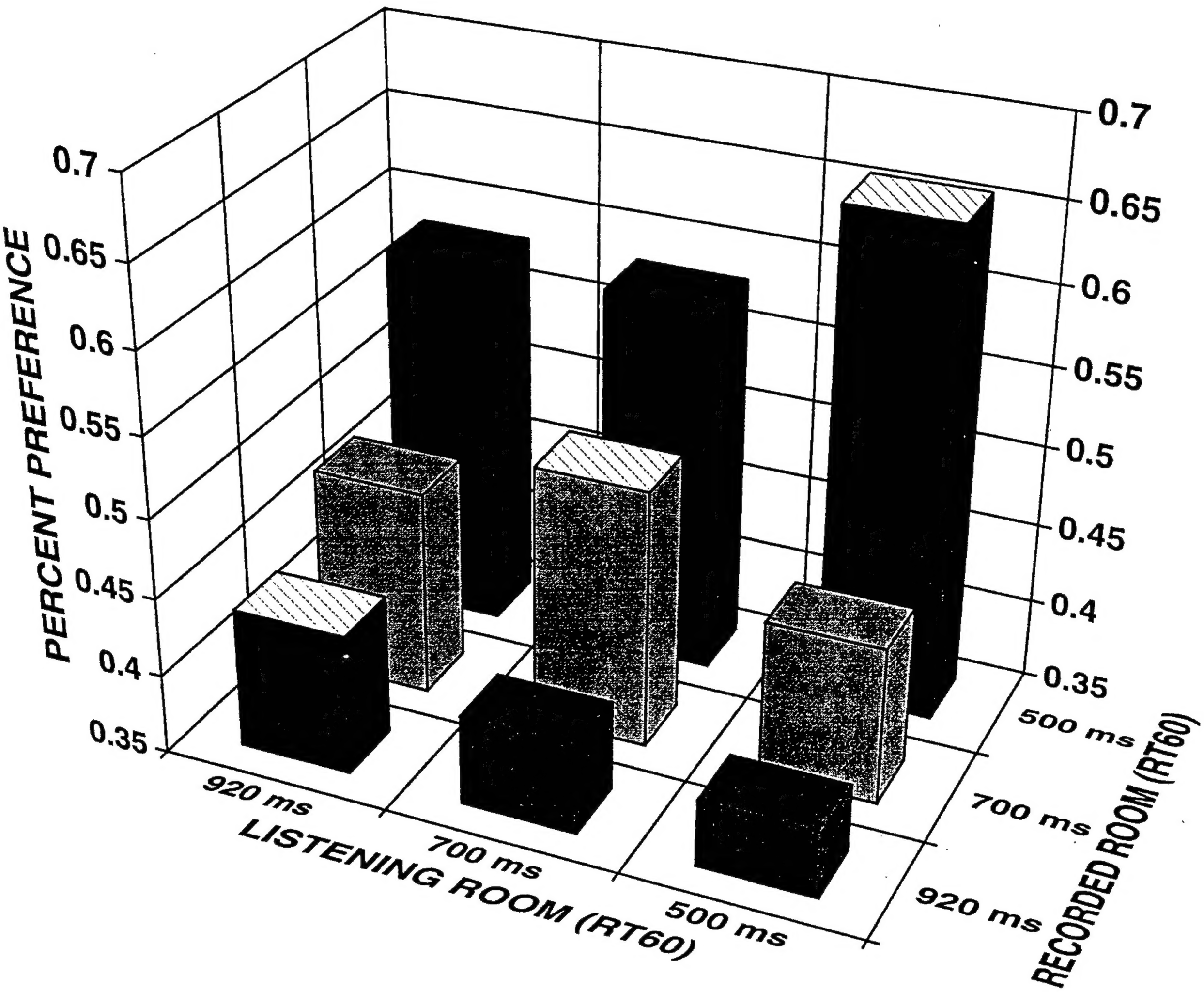
PERCENT PREFERENCE
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
920 ms
700 ms
500 ms
LISTENING ROOM (RT60)
500 ms
700 ms
920 ms
RECORDED ROOM (RT60)

Figure 1

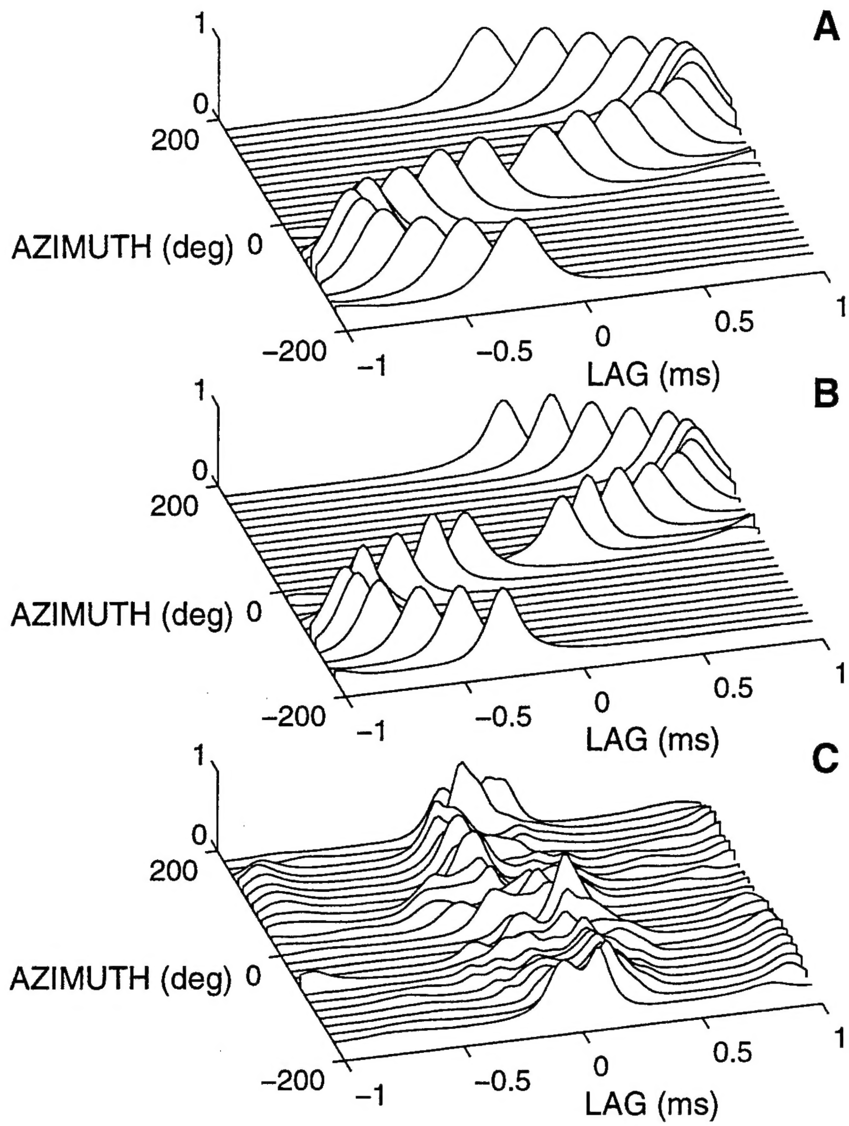
A
1
0
200
AZIMUTH (deg) 0
−200
−1
−0.5
0
0.5
1
LAG (ms)
B
1
0
200
AZIMUTH (deg) 0
−200
−1
−0.5
0
0.5
1
LAG (ms)
C
1
0
200
AZIMUTH (deg) 0
−200
−1
−0.5
0
0.5
1
LAG (ms)

Figure 2

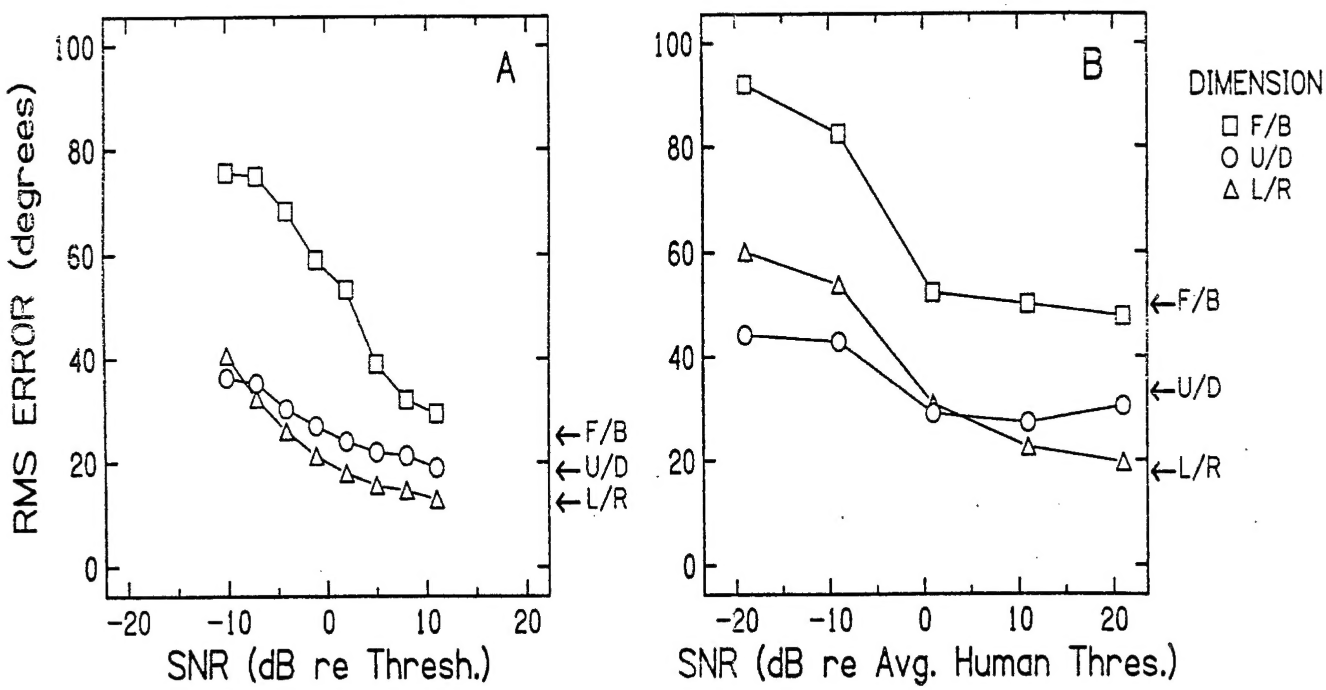
A
B
RMS ERROR (degrees)
100
80
60
40
20
0
-20
-10
0
10
20
SNR (dB re Thresh.)
SNR (dB re Avg. Human Thres.)
←F/B
←U/D
←L/R
DIMENSION
□ F/B
O U/D
Δ L/R

Figure 3

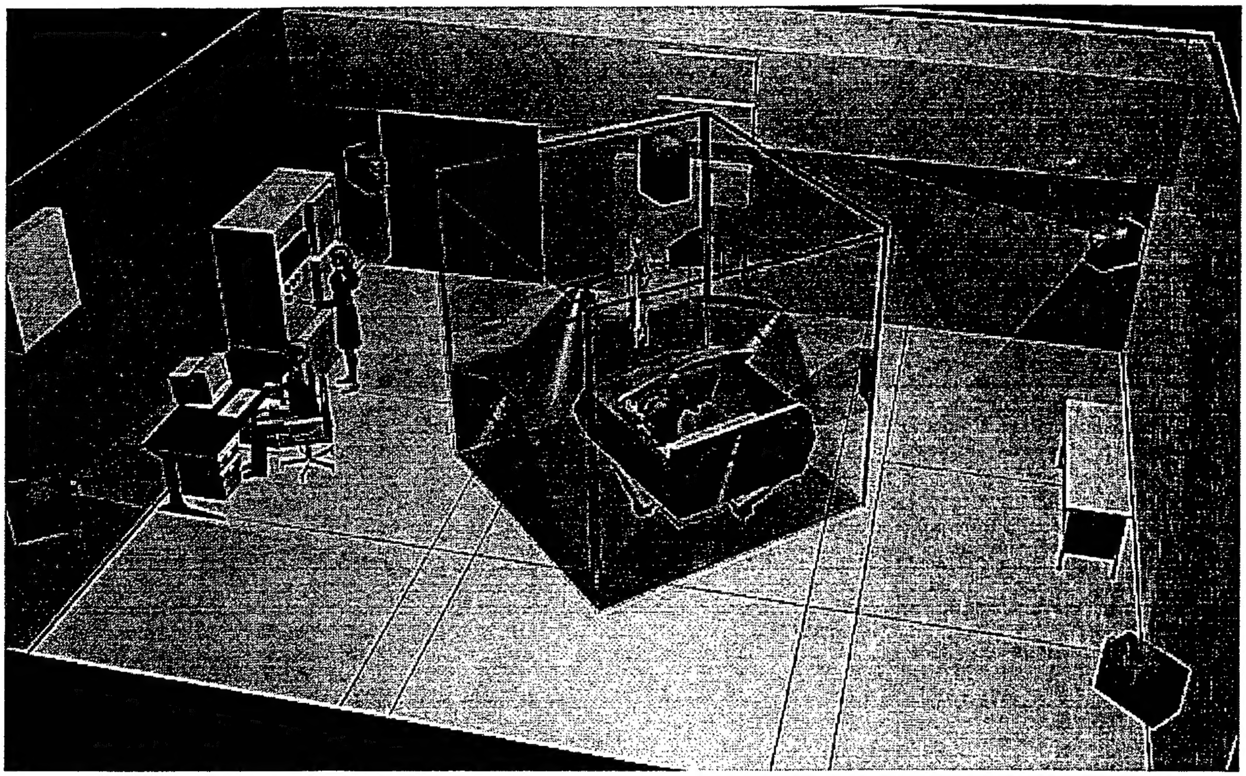

Figure 4